Technical Note 1969-41

E. J. Kelly
R. T. Lacoss

Estimation of Seismicity and Network Detection Capability

16 September 1969

Prepared for the Advanced Research Projects Agency
under Electronic Systems Division Contract AF 19(628)-5167 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ESTIMATION OF SEISMICITY AND NETWORK DETECTION CAPABILITY

E. J. KELLY
R. T. LACOSS

Group 22

TECHNICAL NOTE 1969-41

16 SEPTEMBER 1969

LEXINGTON MASSACHUSETTS

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This research is a part of Project Vela Uniform, which is sponsored by the Advanced Research Projects Agency of the Department of Defense under Air Force Contract AF 19(628)-5167 (ARPA Order 512).

ABSTRACT

The problem of estimating seismicity and the performance of a system which detects earthquakes is formulated in such a way that maximum likelihood estimation can be applied. The mean number of earthquakes which occur in a fixed time interval is assumed to be of the form $\exp\left[a - bm\right]$ where m is magnitude and a and b are constants. The probability of detection as a function of m is taken to be an error function with mean and variance μ and σ. Procedures to obtain maximum likelihood estimates of a, b, μ and σ are derived, discussed, and applied to experimental data to check the relevance of the theoretical development.

Accepted for the Air Force
Franklin C. Hudson
Chief, Lincoln Laboratory Office

TABLE OF CONTENTS

I. INTRODUCTION

Suppose that a seismic station or a network of stations is operated under a fixed set of rules for the detection of earthquakes for a period of time of duration T. Suppose also that only events which prove to be earthquakes from a preassigned region are retained and their body-wave magnitudes determined. The data from such an experiment is an integer, K, the total number of earthquakes recorded, and a list of K magnitudes, $m_1, \ldots, m_K$, usually presented in the form of a cumulative histogram, showing the log of the number of recorded events having magnitude at least m, versus m. Using the available data it should be possible to determine a) the seismicity of the region in question, or b) the detection capability of the station or network used, or c) both. In fact, one must almost always determine both seismicity and detection performance, although the chief objective may be to find only one of these quantities, since the other is rarely well enough known in advance to be modeled accurately.

The aim of this note is to obtain simple rules for the estimation of seismicity and system performance from the data of such an experiment, and to get some idea of the statistical validity of these estimates. In order to reduce this problem to the estimation of specific parameters from the data (so that standard statistical estimation techniques may be applied), it is necessary to make specific assumptions about the random character of

seismicity and the detection of weak events in background noise by the seismic network. In order to push the mathematics through we make use of the well-known and tractable laws of Poisson and Gauss, and our results are only as valid as our underlying assumptions, although the method is quite general.

Specifically, we assume that earthquakes occur as a Poisson process, with rate depending upon magnitude and other parameters. Since in the type of experiment of interest here one is concerned only with total numbers of events which occur over a fixed period of time, we need make no claim that the Poisson process accurately represents the occurrence of events as a stochastic process, but only that the numbers of events recorded having magnitudes (and other parameters) in given ranges are Poisson variables. Let $\bar{N}_m$ be the mean number of events which occur during T having magnitude at least m. In addition to the Poisson assumption further we assume a linear relationship,

$$\log \bar{N}_m = a - bm \quad , \tag{1}$$

between $\log \bar{N}_m$ (natural logarithm) and magnitude. If base-ten logarithms are used, a and b become $a' = a/\log 10$ and $b' = b/\log 10$. Finally, the probability $\Pi(m)$, that the seismic system will detect an event having magnitude m (which will be more precisely defined in Section III) is assumed

to be an error function:

$$\Pi(m) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{m} \exp - \frac{(m'-\mu)^2}{2\sigma^2} dm' \qquad (2)$$

In terms of these models, the general problem is the estimation of the four parameters a, b, μ, and σ from the K + 1 data values K, $m_1, \ldots, m_K$. The quantities a and b define the seismicity of the region, and μ and σ the detection capabilities of the system. In Section II, under the simplifying assumption that $\sigma = 0$ and that μ is known, we discuss some of the implications of the Poisson description of natural seismicity. In Section 3 we formulate and solve the estimation problem in the general case, and in Section IV we illustrate the technique by some application to real data.

II. ESTIMATING SEISMICITY

We assume that an earthquake may be described by its magnitude, m, and a finite set of other parameters, symbolized by the vector $\underline{\alpha}$, with sufficient accuracy so that the a priori probability of its detection by a given station or network may be determined if m and $\underline{\alpha}$ are known. The network must be specified in considerable detail, of course, for such a probability function to be meaningful. This specification will include the nature of the sensors and recording equipment used, station noise levels (and probability distributions) as well as the multi-station data processing algorithm which is used for event detection. The earthquake parameters, $\underline{\alpha}$, will include latitude and longitude of epicenter, depth, fault plane orientation parameters, etc. At the very least, knowledge of the network and specification of the parameters m and $\underline{\alpha}$ should permit the computation of peak signal-to-noise ratios at each of the stations of the system.

Suppose that A stands for a set of possible values of the (vector) parameter $\underline{\alpha}$, and that A_∞ is the set of $\underline{\alpha}$-values which might exist for the region under study. The Poisson "rate function" for events of magnitude m and parameter value $\underline{\alpha}$ will be denoted by $\nu(m, \underline{\alpha})$, and we assume that the probability of occurrence of an earthquake, in any infinitesimal interval dt, having a magnitude in the range m_1 to m_2 and $\underline{\alpha}$-parameter in some set A is

$$\int_{m_1}^{m_2} \int_A \nu(m, \underline{\alpha})\, dm\, d\underline{\alpha}\, dt \qquad (3)$$

Furthermore, according to the Poisson hypothesis, the occurrences which take place in non-overlapping time intervals, or with different magnitudes or α-values, are independent events.

The rate of occurrence of events specified by magnitude alone will be called $\nu(m)$:

$$\nu(m) = \int_{A_\infty} \nu(m, \underline{\alpha}) \, d\, \underline{\alpha} \tag{4}$$

Thus, the probability of occurrence of an event, having magnitude at least m, in time dt is

$$\int_m^\infty \nu(m') \, dm' \, dt$$

and the expected number of such events in a finite interval of duration T is called $\bar{N}_m$:

$$\bar{N}_m = T \int_m^\infty \nu(m') \, dm' \ . \tag{5}$$

We assume that $\bar{N}_m$ has the form

$$\bar{N}_m = e^{a-bm} \ , \tag{6}$$

whence, by differentiation, we obtain

$$\nu(m)\, T = b\, e^{a-bm} \ . \tag{7}$$

The number of events, having magnitudes in the range $m_1 \leq m \leq m_2$, which actually occur during T is a random variable, say, N_{12}. The

expected value of N_{12} is

$$\bar{N}_{12} = T \int_{m_1}^{m_2} \nu(m)\, dm = e^{a} \left(e^{-bm_1} - e^{-bm_2} \right) \tag{8}$$

and the probability that N_{12} actually has the value k is given by the Poisson distribution:

$$\text{Prob} \left\{ N_{12} = k \right\} = \frac{\left(\bar{N}_{12}\right)^k}{k!} e^{-\bar{N}_{12}} \tag{9}$$

Equation (9) is the result of our basic assumptions, and it also follows that if we keep track of the numbers of events which occur during T having magnitudes in a whole set of non-overlapping magnitude intervals, then these numbers are independent random variables, each having a distribution of the form (9) with an appropriate mean value. From this fact it is possible to compute the conditional probability that N_{12} has the value k, given that the total number of events having magnitude at least μ (call this random variable N_μ) is K, where $K \geq k$ and μ is smaller than m_1. This probability is

$$\text{Prob} \left\{ N_{12} = k \,\middle|\, N_\mu = K \right\} = \binom{K}{k} \frac{\left(\bar{N}_\mu - \bar{N}_{12}\right)^{K-k} \left(\bar{N}_{12}\right)^k}{\left(\bar{N}_\mu\right)^K} \tag{10}$$

According to (6), $\bar{N}_\mu = \exp\,(a - b\mu)$.

From the probability distribution (10), which is simply the familiar binomial distribution, giving the chance of k successes out of K trials, with a success probability of $\bar{N}_{12}/\bar{N}_{\mu}$, we can easily compute the conditional mean and variance of N_{12}, given that $N_{\mu} = K$. The conditional mean, $E_K(N_{12})$ is

$$E_K(N_{12}) = \sum_{k=0}^{K} k \text{ Prob} \left\{ N_{12} = k \mid N_{\mu} = K \right\} = K \frac{\bar{N}_{12}}{\bar{N}_{\mu}} , \tag{11}$$

and the standard deviation, $\sigma_K(N_{12})$ (given that $N_{\mu} = K$), is

$$\sigma_K(N_{12}) = \left\{ K \frac{\bar{N}_{12}}{\bar{N}_{\mu}} \left(1 - \frac{\bar{N}_{12}}{\bar{N}_{\mu}} \right) \right\}^{1/2} \tag{12}$$

We illustrate some implications of the formulas by discussing the interpretation of a hypothetical experiment designed to study seismicity for large events. We assume that one is interested only in the rate of occurrence of clearly detectable events and that he is willing to assume that all events of magnitude at least μ are recorded by the system. The detection-system parameters are then removed from the problem and it remains to estimate the seismicity parameters, a and b, from the data. The remainder of this section is devoted to this problem, and the more general case in which the detection capability of the network is also modeled is treated in later sections. As before, we assume that K events are recorded, with

magnitudes m_i, $i = 1, \ldots, K$, all at least equal to μ. Let N_m stand for the number of recorded events having magnitude at least m. Then $\bar{N}_m$, the unconditioned mean, is $\bar{N}_m = \exp(a-bm)$, and the conditional mean, given that $N_\mu = K$, is given by (11):

$$E_K(N_m) = K\frac{\bar{N}_m}{\bar{N}_\mu} = K\,e^{-b(m-\mu)} \tag{13}$$

The usual display of such experimental data is a plot of $\log N_m$ versus m. In such a plot, $\log(E_K(N_m))$ will appear as a straight line of slope (-b). If base-ten logarithms are used, we rewrite (13) as

$$E_K(N_m) = K\,10^{-b'(m-\mu)} \tag{14}$$

where $b' = b/\log 10 = 0.43429\ldots b$.

As m increases, the expected number of events decreases rapidly, and with it, the ratio of standard deviation (formula (12)) to mean increases. In fact, for very small values of $E_K(N_m)$, $\sigma_K(N_m) \approx (E_K(N_m))^{1/2}$. Since the conditional probability distribution for N_m (given $N_\mu = K$) is the binomial law (10), we can establish intervals, about the value of $E_K(N_m)$, which contain the value of N_m with fixed probability. This is illustrated in Figures 1 and 2 for K = 100 and K = 1000, where the solid lines are plots of $E_K(N_m)$ versus m (assuming that $b' = 1.0$) and the dashed lines contain the

values of N_m with 50% probability (i.e., they mark the points outside of which the tails of the distribution contain 25% probability) and 90% probability (i.e., 5% tails), as labeled.

These illustrations show that it is difficult to fit a straight line to data, presented graphically in this form, and put the proper statistical weight to the various portions of the curve. The determination of a straight line fit by simple least-squares has also been criticized on this basis (2). The problem is seriously aggravated in the usual case, where the cumulative histogram begins to level off, going towards lower magnitudes, because the detection system begins to miss events, just as the number of recorded events is beginning to give a statistically reliable picture of the seismicity law. For these reasons, statistical estimation theory is applied to this problem, using the maximum likelihood method for the determination of unknown parameters.

From the independence of the numbers of events in different magnitude ranges, it follows that the conditional probability, given $N_\mu = K$, that the K recorded events actually have magnitudes in the intervals m_i to $m_i + dm_i$, $i = 1, \ldots K$, is

$$K! \prod_{i=1}^{K} \frac{\nu(m_i)\, T\, dm_i}{\bar{N}_\mu} = K! b^K e^{b\mu K} \prod_{i=1}^{K} e^{-bm_i} dm_i$$

$$\equiv P_K(m_1 \ldots m_K)\, dm_1 \ldots dm_K \tag{15}$$

We assume that the observed values of magnitude, $m_1 \ldots m_K$, are arranged in non-decreasing order (not in the order of occurrence). Thus $P_K(m_1 \ldots m_K)$ is a probability density function, defined and normalized over the portion of K-space given by

$$m_1 \leq m_2 \leq \ldots \leq m_K \quad .$$

The corresponding likelihood function, Λ_K is given by log P_K, and the maximum likelihood estimator of b is the value which maximizes Λ_K. We denote by $\langle m \rangle$ the average observed magnitude:

$$\langle m \rangle \equiv \frac{1}{K} \sum_{l=1}^{K} m_i \quad , \tag{16}$$

and we compute

$$\Lambda_K(m_1, \ldots, m_K \,;\, b) = K \left[\log b + (\mu - \langle m \rangle)\, b \right] + \log K!$$

The value, $\hat{b}$, which maximizes Λ_K is obviously

$$\hat{b} = (\langle m \rangle - \mu)^{-1} \quad . \tag{17}$$

Because of the form of (17), it is more convenient to parametrize the seismicity by an inverse parameter, m_e:

$$m_e \equiv 1/b \tag{18}$$

so that the mean seismicity curve has the form

$$E_K(N_m) = e^{a-(m/m_e)} \tag{19}$$

Then the estimator of m_e corresponding to (17) is

$$\hat{m}_e = \langle m \rangle - \mu \quad . \tag{20}$$

Equation (20) implies that the seismicity parameter, m_e, is chosen so that the expected number of events of magnitude at least equal to the sample mean $\langle m \rangle$ is just (1/e) times the number actually contained in the sample (which contains only events of magnitude at least μ). If one had a really large sample of events covering a large range of magnitudes, it would be possible to test the assumption of exponential seismicity, as expressed by (19), by treating the data as a series of experiments, each containing the data of its predecessors, and each having a smaller value of μ.

One can get some idea of the accuracy of the estimator (20) by computing its mean and standard deviation, conditioned on $N_\mu = K$. By direct integration, using the conditional probability density (15) we find that

$$E_K(\langle m \rangle) = b\, e^{b\mu} \int_\mu^\infty e^{-bm}\, m\, dm = \mu + 1/b$$

or

$$E_K(\hat{m}_e) = 1/b = m_e \quad . \tag{21}$$

In other words, our estimator of m_e is unbiased. We also find

$$E_K(\langle m \rangle^2) = (\mu + m_e)^2 + (m_e^2/K) \quad ,$$

so that the standard deviation of $\hat{m}_e$ (given $N_\mu = K$) is

$$\sigma_K(\hat{m}_e) = m_e/K^{1/2} \quad . \tag{22}$$

If we put $b' = 1/m_d$, so that (19) is the same as

$$E_K(N_m) = 10^{a'-(m/m_d)} \quad ,$$

then $m_d = (\log 10)m_e$ and $\hat{m}_d = (\log 10) \cdot \hat{m}_e$. Both estimators have the same relative accuracy:

$$\frac{\sigma_K(\hat{m}_e)}{E_K(\hat{m}_d)} = \frac{\sigma_K(\hat{m}_d)}{E_K(\hat{m}_d)} = K^{-1/2} \tag{23}$$

Thus, a 100-earthquake sample should provide 10% accuracy in the measurement of seismicity.

So far we have ignored the parameter a, since it drops out of the conditional probability density from which b was estimated. But the fact that K events were seen, with magnitude at least μ, is itself a measure of total seismicity, and hence can be used to determine a. According to our basic model, the probability that N_μ is equal to K is given by (9):

$$\text{Prob}\left\{N_\mu = K\right\} = \frac{(\bar{N}_\mu)^K}{K!} \, e^{-(\bar{N}_\mu)} \quad , \tag{24}$$

where $\bar{N}_\mu = \exp(a-\hat{b}\mu)$, and $\hat{b}$ has already been estimated. The maximum likelihood estimator of a is that value, $\hat{a}$, which maximizes (24), for fixed K. But the function $X^K \exp(-X)$ is a maximum at $X = K$, hence the

estimate, $\hat{a}$, must make $\bar{N}_\mu = K$, i.e.,

$$\hat{a} = \hat{b}_\mu + \log K \quad . \tag{25}$$

If we approach the problem in terms of the joint estimation of a and b, we are led immediately to the same results, as will be seen in the next section, when a more general problem is solved.

It should be observed that equation (15) lends itself to a slightly different interpretation. Instead of thinking of a single experiment which yields the magnitude values $m_1, m_2, \ldots, m_K$ (listed in non-decreasing order), we may think of the result as the outcome of K independent trials of a single experiment, each of which yields a single value of magnitude. The probability density for magnitude, now treated as a random variable, is

$$f(m) = \frac{\nu(m)T}{\bar{N}_\mu} = be^{-b(m-\mu)}$$

which is normalized in the interval $\mu \leq \infty$. If $\tilde{m}_i$ is the magnitude of the i^{th} event, then the probability density for the sequence $\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_K$ is, of course,

$$\prod_{i=1}^{K} f(\tilde{m}_i) \equiv F_K(\tilde{m}_1, \ldots, \tilde{m}_K)$$

Finally, the probability density for obtaining the values $m_1, m_2, \ldots, m_K$ (listed in non-decreasing order) will be the sum over all possible orders at occurrence, i.e., $K!\ F_K(m_1, \ldots, m_K)$, which is exactly $P_K(m_1, \ldots, m_K)$,

as given by (15). In the literature (1)(2)(4)(5)(6) the problem has been treated from this second point of view, starting with $f(m)$ and $F(m_1, \ldots, m_K)$ as the basic probability laws. In this way the estimator (17) has been found, and formulas (21) and (22) derived. In fact, the probability density of $\hat{b}$ has been obtained and confidence intervals established for this estimator by Aki and Utsu.

III. ESTIMATING NETWORK PERFORMANCE

In Section II we introduced the "rate function," $\nu(m, \underline{\alpha})$, which describes the relative rate of occurrence of earthquakes in terms of magnitude and the vector parameter $\underline{\alpha}$. The ratio of this rate to the integrated rate $\nu(m)$ will be denoted by

$$f(\underline{\alpha} | m) \equiv \frac{\nu(m, \underline{\alpha})}{\nu(m)} \tag{26}$$

Because of the meaning of $\nu(m, \alpha)$ dm d$\underline{\alpha}$ dt and $\nu(m)$ dm dt as probabilities, $f(\alpha | m)$ itself is the conditional probability density that, given the occurrence of an event with magnitude m, the event had a parameter-value $\underline{\alpha}$. We have assumed that $\underline{\alpha}$ and m together characterize an event well enough so that its probability of detection by the network can be specified. Let this detection probability be $P(m, \alpha)$. Then the probability that, in time dt, an event in the magnitude range m to m + dm will occur and be detected is

$$\int_{A_\infty} P(m, \underline{\alpha})\, \nu(m, \underline{\alpha})\, d\underline{\alpha}\, dm\, dt$$

$$= \nu(m)\, dm\, dt \int_{A_\infty} P(m, \underline{\alpha})\, f(\underline{\alpha} | m)\, d\underline{\alpha} \quad .$$

We define the integral which appears here to be $\Pi(m)$:

$$\Pi(m) \equiv \int_{A_\infty} P(m, \underline{\alpha})\, f(\underline{\alpha} | m)\, d\underline{\alpha} \tag{27}$$

This quantity is clearly the probability that the network detects an event of magnitude m, knowing only m, but taking account of the different ways

this can happen (different $\underline{\alpha}$-values) and weighting them according to the detailed seismicity of the region, as described by $f(\underline{\alpha}|m)$. In general, we do not know $f(\underline{\alpha}|m)$, and cannot readily estimate it from the data, hence we must work directly with $\Pi(m)$, the effective detection probability, and we will make simple assumptions about the form of this function, in spite of its dependence on the nature of the region studies. In terms of $\Pi(m)$, the rate of recorded events over an interval of duration T, having magnitudes in the range $m_1 \leq m \leq m_2$, is

$$\bar{N}_{12} = T \int_{m_1}^{m_2} \Pi(m)\, \nu(m)\, dm = be^{a} \int_{m_1}^{m_2} \Pi(m)\, e^{-bm} dm \tag{28}$$

The effective detection probability, $\Pi(m)$, must contain parameters characteristic of the system which can be estimated from the data. For one of these parameters we choose the magnitude, μ, for which the effective detection probability is 50%, i.e., we put

$$\Pi(m) = \Pi_o(m-\mu) \tag{29}$$

where $\Pi_o(m)$ is a function which increases from zero (at $m \to -\infty$) to unity ($m \to +\infty$), and goes through the value 1/2 at m = 0. Later, we shall specialize $\Pi_o(m)$ to be an error function, but it is convenient to carry out the analysis using form (29) as far as the estimation of a, b, and μ. At that point we shall introduce the error-function hypothesis for $\Pi_o(m)$ (the reader may be moved to try other models) and estimate also the variance

parameter, σ (see equation 2).

Thus, the expected number of recorded events in the magnitude range $m_1 \leq m \leq m_2$ (formula 28) becomes

$$N_{12} = be^{a-b\mu} \int_{m_1-\mu}^{m_2-\mu} \Pi_o(m)\, e^{-bm}\, dm \tag{30}$$

From (30), we see that the expected number of recorded events of magnitude at least m is

$$\overline{N}_m = b\, e^{a-b\mu} \int_{m-\mu}^{\infty} \Pi_o(m')\, e^{-bm'}\, dm' \tag{31}$$

and the expected total number of recorded events is

$$\overline{N} = b\, e^{a-b\mu}\, I(b) \quad , \tag{32}$$

where

$$I(b) \equiv \int_{-\infty}^{\infty} \Pi_o(m)\, e^{-bm}\, dm \quad . \tag{33}$$

It is assumed that $\Pi_o(m)$ vanishes rapidly enough, as $m \to -\infty$, to guarantee the convergence of this integral.

We can now compute the likelihood function for the problem and formulate the estimation of all the unknown parameters. As before, the experiment consists of observation of a given region for a time of duration T, using a fixed detection network, and this time K represents the total number of events recorded, regardless of magnitude, while the actual magnitudes observed are m_i, $i = 1, \ldots, K$. For a given set of parameters,

a, b, μ (and others implicit in $\Pi_o(m)$), the probability of recording just K events in time T is the Poisson law

$$\text{Prob}\left\{N = K\right\} = \frac{(\bar{N})^K}{K!} e^{-\bar{N}} \quad , \tag{34}$$

with $\bar{N}$ given by (32). If we define K non-overlapping magnitude intervals and if $\bar{N}_i$ is the expected number of events in the i^{th} interval (given by formula (30) with appropriate values of m_1 and m_2), then the conditional probability, given K detections, of obtaining just one event in each of these K magnitude intervals is

$$K! \prod_{i=1}^{K} (\bar{N}_i/\bar{N}) \quad . \tag{35}$$

When the magnitude intervals are made infinitesimal, so that the i^{th} ranges from m_i to $m_i + dm_i$, we have

$$\bar{N}_i = T\, \Pi(m_i)\, \nu(m_i)\, dm_i$$

and the total probability of recording K events with magnitudes in these K intervals is

$$P(K, m_1, \ldots, m_K)\, dm_1 \ldots dm_K \quad ,$$

where

$$P(K, m_1, \ldots, m_K) = e^{-\bar{N}} \prod_{i=1}^{K} T\, \Pi(m_i)\, \nu(m_i)$$

$$= e^{-\bar{N}}\, b^K\, e^{K\left[a - b\langle m\rangle\right]} \prod_{i=1}^{K} \Pi_o(m_i - \mu) \tag{36}$$

In (36) we have again made use of the symbol $\langle m \rangle$ for the average of the observed magnitudes, as given by 16), and $\overline{N}$ is given explicitly by (32). $P(K, m_1, \ldots, m_K)$ is an absolute probability, not a conditional probability like $P_K(m_1 \ldots m_K)$. The expression $P(K, m_1 \ldots m_K)$ is also defined for magnitude variables satisfying $m_1 \leq m_2 \leq \ldots \leq m_K$. Finally, the likelihood function is

$$\Lambda(K, m_1, \ldots, m_K \,; a, b, \mu) = \log P(K, m_1, \ldots, m_K)$$
$$= -\overline{N} + K\left[a - b\langle m \rangle + \log b\right] + \sum_{i=1}^{K} \log \Pi_o(m_i - \mu) \tag{37}$$

The maximum-likelihood estimators of a, b and μ are obtained by maximizing Λ over these parameters. Differentiating with respect to a, we require that

$$\frac{\partial \Lambda}{\partial a} = -\frac{\partial \overline{N}}{\partial a} + K = 0$$

But, according to (32),

$$\frac{\partial \overline{N}}{\partial a} = \overline{N} \quad ,$$

hence, the estimator, $\hat{a}$, is that value of a for which (with given b, μ) $\overline{N} = K$. This result is already interesting in that it states that the overall seismicity, which if governed by the parameter a, must be chosen so that the expected total number of recorded events is equal to the actual number seen. It is easily seen that (34) is maximized when $\overline{N} = K$, hence estimating the mean of a sample Poisson distribution from a single measurement

produces the same result we have found here for the more complex situation of a Poisson process with parameters.

From (32) we can write

$$\log \overline{N} = a - b\mu + \log b + \log I(b) \quad ,$$

hence, the estimator which equates $\overline{N}$ to K is

$$\hat{a} = \log K + b\mu - \log b - \log I(b) \quad . \tag{38}$$

Substituting this value into (37) we obtain

$$\Lambda(K_1, m_1, \ldots, m_K \,;\, \hat{a},\ b, \mu) = K\left[\log K - 1 + b\mu - b\langle m\rangle - \log I(b)\right] + \sum_{i=1}^{K} \log \Pi_o\,(m_i - \mu) \quad . \tag{39}$$

The equations for the estimation of b and μ are obtained by setting the partial derivatives of Λ_K, given by (39), with respect to these parameters equal to zero:

$$\mu - \langle m\rangle - \frac{I'(b)}{I(b)} = 0 \tag{40a}$$

$$b - \frac{1}{K} \sum_{i=1}^{K} \frac{\Pi_o{}'(m_i - \mu)}{\Pi_o(m_i - \mu)} = 0 \tag{40b}$$

In equations (40) we have used the prime to denote the derivative of a function with respect to its argument. Thus,

$$I'(b) = \frac{dI(b)}{db} = -\int_{-\infty}^{\infty} \Pi_o(m)\, e^{-bm}\, m\, dm \tag{41}$$

(we assume that $\Pi_o(m)$ vanishes rapidly enough, as $m \to -\infty$, so that the first several derivatives of I(b) exist).

In order to proceed, we must now make an assumption for $\Pi_o(m)$, perhaps containing further parameters. We shall then have explicit equations for μ and b, and, by differentiating Λ with respect to the new parameters, we get equations for them as well. In passing we note that the case treated in Section II can be retrieved by treating μ as known (hence ignoring (40b)) and putting $\Pi_o(m)$ equal to zero for negative m, unity for positive m. We also assume that each $m_i \leq \mu$, since the contrary would be contradictory and would also make P vanish. Then I(b) becomes simply 1/b, and (40a) immediately reproduces (17).

As promised in the Introduction, we shall assume an error-function law for $\Pi_o(m)$. We shall ultimately show by example in Section IV that it can provide a reasonable representation of experimental results. Define

$$\mathrm{erf}(x) \equiv (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-y^2/2)\, dy \tag{42}$$

and take

$$\Pi_o(m, \sigma) = \mathrm{erf}(m/\sigma) = 2\pi\sigma^2)^{-1/2} \int_{-\infty}^{m} \exp(-y^2/2\sigma^2)\, dy \quad . \tag{43}$$

With this assumption we find (integrating by parts)

$$I(b) = I(b, \sigma) = \int_{-\infty}^{\infty} \Pi_o(m, \sigma)\, e^{-bm}\, dm = \frac{1}{b} \int_{-\infty}^{\infty} e^{-bm} \frac{\partial \Pi_o(m, \sigma)}{\partial m}\, dm \quad .$$

But

$$\frac{\partial \Pi_o(m,\sigma)}{\partial m} = (2\pi\sigma^2)^{-1/2} \exp(-m^2/2\sigma) \quad ,$$

the Gauss function, hence,

$$I(b,\sigma) = (2\pi b^2\sigma^2)^{-1/2} \int_{-\infty}^{\infty} e^{-\left[\frac{m^2}{2\sigma^2} + bm\right]} dm = \frac{1}{b}\, e^{\frac{b^2\sigma^2}{2}} \tag{44}$$

It follows from this that

$$I'(b) = (\sigma^2 - \frac{1}{b^2})\, e^{\frac{b^2\sigma^2}{2}}$$

Using the error function for Π_o, equation (40a)) for the determination of $\hat{b}$ is simply

$$\hat{b}\sigma^2 - \frac{1}{\hat{b}} = \mu - \langle m \rangle \quad . \tag{45a}$$

Again, we replace b by its inverse, m_e, which parameter is estimated by $\hat{m}_e$, determined from

$$\hat{m}_e - \frac{\sigma^2}{\hat{m}_e} = \langle m \rangle - \mu \quad . \tag{45b}$$

In the limit $\sigma \to 0$, our function $\Pi_o(m,\sigma)$ reverts to the simple zero-or-one probability which describes the case treated in Section II and (45b) obviously reduces to equation (20) of that section. In general, the solution of (45b), which reduces to the proper limit as $\sigma \to 0$, is

$$\hat{m}_e = \frac{1}{2}(\langle m \rangle - \mu)\left[1 + \sqrt{1 + \frac{4\sigma^2}{(\langle m \rangle - \mu)^2}}\right]$$

$$= \langle m \rangle - \mu + \frac{\sigma^2}{\langle m \rangle - \mu} - \frac{\sigma^4}{(\langle m \rangle - \mu)^3} + \ldots \qquad (46)$$

In expression (46) the parameters of the detection system, μ and σ, are yet to be estimated. When their estimates are found, they are substituted in (46) and (38) to provide the final estimates of the seismicity parameters. However, if one felt that the detection system were adequately represented by $\Pi(m)$, as given by (29) and (43) with known values of μ and σ, then (38) and (46) would be final estimates of a and m_e. It is logical in that case to ask about the statistical properties of $\hat{m}_e$ as an estimator of m_e. However, the form of (46) makes a direct evaluation of the conditional mean and variance of $\hat{m}_e$ impossible. The best we can do is evaluate the mean and variance of $\langle m \rangle$, as we did in Section II in a simpler problem, and relate these quantities indirectly to $\hat{m}_e$, or $\hat{b}$, through (40a). These calculations are similar to those leading to (21) and (22), and we obtain

$$E_K(\langle m \rangle) = \frac{b\, e^a \int_{-\infty}^{\infty} e^{-bm} \Pi_o(m-\mu) m \, dm}{b\, e^a \int_{-\infty}^{\infty} e^{-bm} \Pi_o(m-\mu) \, dm} = \frac{\mu\, I(b) - I'(b)}{I\,(b)} = \mu - \frac{I'(b)}{I(b)}$$

and

$$E_K(\langle m\rangle^2) = (E_K\langle m\rangle)^2 + \frac{1}{K}\left\{\frac{I''(b)}{I(b)} - \left[\frac{I'(b)}{I(b)}\right]^2\right\}$$

Thus

$$E_K(\langle m\rangle - \mu) = -\frac{I'(b)}{I(b)} \tag{47a}$$

and

$$\sigma_K(\langle m\rangle - \mu) = K^{-1/2}\left\{\frac{I''(b)}{I(b)} - \left[\frac{I'(b)}{I(b)}\right]^2\right\}^{1/2} \tag{47b}$$

Let us put

$$\Phi(b) \equiv \frac{I'(b)}{I(b)} = \frac{d}{db}\log I(b) \tag{48}$$

Then (40a) says that the estimator, $\hat{b}$, is determined by

$$-\Phi(\hat{b}) = \langle m\rangle - \mu \quad ,$$

and (47a) says, in turn, that

$$E_K\,(\Phi(\hat{b})) = \Phi(b) \quad , \tag{49a}$$

i.e., $\Phi(\hat{b})$ is an unbiased estimator of $\Phi(b)$, if μ and σ are known. Moreover, equation (47b) is equivalent to

$$\sigma_K\,(\Phi(\hat{b})) = \left[\Phi'(b)/K\right]^{1/2} \tag{49b}$$

Equations (49) together show that, as $K \to \infty$, the "estimator" $\Phi(\hat{b})$ converges (in the mean) to $\Phi(b)$, and in this sense $\hat{b}$ must become a better and better estimate of b.

To continue with the general problem, we now have explicit formulas for $\hat{a}$ and $\hat{b}$ in terms of μ and σ. In order to determine μ and σ we must

minimize $\Lambda(K, m_1, \ldots, m_K ; \hat{a}, \hat{b}, \mu, \sigma)$ over these parameters. Since equation (40b), equivalent to $\partial\Lambda/\partial\mu = 0$, and a similar equation, $\partial\Lambda/\partial\sigma = 0$, cannot be solved explicitly, one has to proceed numerically, and for this purpose it is simpler to work with the likelihood function itself, seeking a minimum by some search procedure. According to (45a),

$$\hat{b}(\mu - \langle m \rangle) = (\hat{b}\sigma)^2 - 1$$

and also, from (44)

$$\log I(\hat{b}) = \frac{1}{2}(\hat{b}\sigma)^2 - \log \hat{b} \quad .$$

When these expressions are used in formula (39) we obtain the desired expression for the likelihood function:

$$\Lambda(K, m_1, \ldots, m_K ; \hat{a}, \hat{b}, \mu, \sigma) = K\left[\log K - 2 + \frac{1}{2}(\hat{b}\sigma)^2 + \log \hat{b}\right] + \sum_{i=1}^{K} \log \Pi_o (m_i - \mu, \sigma) \tag{50}$$

In (50), of course, $\hat{b} = 1/\hat{m}_e$, with $\hat{m}_e$ given by (46). In the next section we give a numerical example of the use of (50) and (46) to estimate μ and σ as well as b.

A slight change in viewpoint to that described at the end of Section II allows one to invoke general properties of well-behaved maximum likelihood estimators. Specifically, as $K \to \infty$ our experimental data can be considered as resulting from larger and larger numbers of independent samples from

a single distribution with unknown parameters. As such, the maximum likelihood estimators of these parameters (μ, σ, b) are consistent and have minimum variance. This is, of course, a much stronger statement than our previous observation that $\Phi(\hat{b}) \rightarrow \Phi(b)$ if μ and σ are known.

Before we proceed with numerical examples, let us see what are some further implications of our assumed form for $\Pi(m)$ concerning the outcome of a seismicity experiment. In particular, treating all the parameters as fixed, let us compute the expected number of recorded events, $\bar{N}_{12}$, in a magnitude interval $m_1 \leq m \leq m_2$. Just as in Section II, the actual number, N_{12}, is a Poisson variable (i.e., (9) is valid), and the numbers of events recorded in a set of non-overlapping magnitude intervals are independent random variables. From (30), on integrating by parts,

$$\begin{aligned}
\bar{N}_{12} &= -e^{a-b\mu}\left[\Pi_o(m_2-\mu,\sigma)\, e^{-b(m_2-\mu)} - \Pi_o(m_1-\mu,\sigma)\, e^{-b(m_1-\mu)}\right] \\
&\quad + e^{a-b\mu}\int_{m_1-\mu}^{m_2-\mu} e^{-bm}\, \Pi_o'(m,\sigma)\, dm \\
&= e^{a-bm_1}\, \mathrm{erf}\left(\frac{m_1-\mu}{\sigma}\right) - e^{a-bm_2}\, \mathrm{erf}\left(\frac{m_2-\mu}{\sigma}\right) \\
&\quad + e^{a-b\mu+\frac{b^2\sigma^2}{2}}\left[\mathrm{erf}\left(\frac{m_2-\mu+b\sigma^2}{\sigma}\right) - \mathrm{erf}\left(\frac{m_1-\mu+b\sigma^2}{\sigma}\right)\right]
\end{aligned} \qquad (51)$$

By putting $m_1 = m$, $m_2 = +\infty$, we obtain the expected number of recorded events having magnitude at least m:

$$\bar{N}_m = e^{a-bm} \operatorname{erf}\left(\frac{m-\mu}{\sigma}\right) + e^{a-b\mu+\frac{b^2\sigma^2}{2}} \left[1 - \operatorname{erf}\left(\frac{m-\mu+b\sigma^2}{\sigma}\right)\right], \tag{52}$$

and the expected total number of events:

$$\bar{N} = \bar{N}_{-\infty} = e^{a-b\mu+\frac{b^2\sigma^2}{2}} \tag{53}$$

Given that an experiment produced a total of K recorded events (of all magnitudes) then the probability that N_m, the number having magnitude at least m, is equal to k is given (just as in (10)) by the binomial distribution

$$\text{Prob}\left\{N_m = k \mid N = K\right\} = \binom{K}{k} p^k q^{K-k} \tag{54}$$

where $q \equiv 1 - p$ and the "success probability," p, is

$$p = \bar{N}_m / \bar{N} \ . \tag{55}$$

The conditional mean of N_m, given that $N = K$ is just like (11):

$$E_K(N_m) = Kp \tag{56}$$

Equations (56) and (54) are used in Section IV to indicate that the model developed, including use of the error function, gives a reasonable picture of observed seismicity and the fluctuations to be expected in the observed number of events.

IV. EXPERIMENTAL RESULTS

We have applied the maximum likelihood method described in the preceding section to samples of worldwide data obtained from the U. S. C. and G. S. Our objective has not been to evaluate the network or to obtain better estimates of worldwide seismicity. Rather, the objective has been to check the relevance to real data of the theoretical machinery which has been developed. It would appear that the maximum likelihood method and the statistical model used are indeed adequate and supply considerable insight into the problem of estimating seismicity.

The first 2000 events of 1968 reported by U. S. C. and G. S. have been used to estimate μ, σ, a' and b'. The primed (base 10) seismicity parameters were estimated to be consistent with standard seismological usage. Constant terms were subtracted from the likelihood equation (50) and the result has been plotted as solid contours on Figures 3a and 3b. Note that the contour intervals are not uniform. The function $\hat{b}' = 1/(\hat{m}_e \log 10)$, with $\hat{m}_e$ given by (46) is also shown by dashed contour lines on the figures. The likelihood function in this example is maximized by $\hat{\mu} = 5.1$, $\hat{\sigma} = 0.415$, and $\hat{b}' = 1.725$. The corresponding value of $\hat{a}$, assuming a one-year time period, has also been computed by solving $\log K = \hat{a} - \hat{b}\hat{\mu} + \frac{\hat{b}^2\hat{\sigma}^2}{2}$ for $\hat{a}$ where K is not 2000 but 4500 since the 2000th event occurred on 11 June 1968. From this we obtained $\hat{a}' = \hat{a} \log 10 = 11.9$. The estimated values of

a' and b' are in reasonable agreement with results obtained by Gutenberg (3) using only large events over an extended time period. The Gutenberg values are $b' = 1.59$ and $a' = 12.2$ if it is assumed that surface wave magnitudes, M_s, are related to m_b by $M_s = 1.59m_b - 3.97$. (3)

Figure 4 shows the experimental data used in the above experiment. The data has been normalized to obtain the percent of events observed above magnitude m_b as a function of m_b. The theoretical success probability $p = \overline{N}_m/\overline{N}$, with $\overline{N}_m$ given by (52) and $\overline{N} = 2000$, has also been shown for $\mu = 5.1$, $\sigma = 0.415$, and $b' = 1.725$ which are the maximum likelihood parameter estimates.

In some applications of the maximum likelihood method of parameter estimation it has been noted that the estimates are also optimum in the sense of some reasonable criterion for fitting data with known parameterized functions and that the likelihood function is a measure of the goodness of fit as a function of parameter values. The most trivial example is the estimation of the mean of a Gaussian distribution using N independent samples from the distribution. In this case the log likelihood function can be taken as $\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$ where the x_i are the observations. Then the maximum likelihood estimator is a least mean square estimator and the likelihood function is just the mean square error.

We have been unable to explicitly demonstrate a simple curve-fitting

interpretation to the likelihood function for the problem considered in this paper. Nevertheless it appears to be qualitatively reasonable to accept the assertion that the likelihood function is a measure of the degree of agreement between experimental data and the parameter values for which the likelihood function is evaluated. For example, consider the μ, σ and b values which have been indicated by the four symbols on the -1327 contour on Figure 3b. A few of the success probability values implied by these parameter values have been indicated on Figure 4. None of these μ, σ, and b' values do significantly more violence to the experimental data than do the maximum likelihood estimates. Thus, from the subjective viewpoint of fitting the experimental success probability, the small perturbation of the likelihood function has had only a small effect. Except for the maximum likelihood estimates no values of success probability have been shown below $m_b = 5.5$ since the different values are not clearly distinguishable on the figure at lower magnitudes. Other values of μ, σ, and b', which correspond to much larger changes of likelihood, have also been checked and found to result in large disagreement between the data and theoretical success probability.

From the point of view of fitting a curve to the experimental data on Figure 4, it is clear that significant changes in μ, σ, and b' from the maximum likelihood values can have only a limited effect if the changes are

coordinated to have little effect on the likelihood function. Thus, moving μ and σ along the ridge of Λ by 10% and making a corresponding change in b' of about 30% has only a limited effect on the value of the likelihood function and the fit of the success probability to experimental data. It should be noted that this discussion does not directly relate to stability of the maximum likelihood estimates. However, one might conjecture that the shape and steepness of the maximum of the likelihood function is closely related to stability. In this case, it would seem that estimates of b' will tend to be less stable than those for μ and σ if all three are to be estimated. In addition, the curvature of the likelihood function in the region of its maximum suggests that satisfactory stability may require long periods of observation.

The stability of estimates is also related to the fluctuations in experimental success probabilities which might be obtained for a typical experiment. Such fluctuations have been investigated using a short run of U.S.C. and G.S. data (90 events from two PDE cards). The short run of data was used to emphasize the fluctuations. The likelihood function showed a very broad minimum, but the parameter values $b' = 1.65$, $\mu = 4.9$, and $\sigma = 0.39$ were near the minimizing set. The curve of the success probability, computed from equations (52) through (56), for these parameter values is plotted as the solid line on Figure 5. The experimental data are also shown

on the figure. By using tables of the cumulative binomial distribution for 90 trials and various success probabilities we determined integers which most nearly satisfied the equations.

$$\sum_{k=0}^{k_5} \text{Prob}\left\{N_m = k / N = K\right\} = 5\%$$

and

$$\sum_{k=k_{95}}^{K} \text{Prob}\left\{N_m = k / N = K\right\} = 5\%$$

In other words, k_5 and k_{95} determine the 5% tails as used in Section II. Using (55) to relate success probability to magnitude and normalizing k_5 and k_{95} as a fraction of K, we obtain the dashed curves in Figure 5. It appears that the model gives a reasonable picture both of the average seismicity and the fluctuations to be expected about the mean.

REFERENCES

(1) Aki, Keiiti, "Maximum Likelihood Estimate of b in the Formula log N = a-bM and its Confidence Limits; Bull. of the Earthquake Research Institute, V43 (1965), pp. 237-239.

(2) Page, Robert, "Aftershocks and Microaftershocks of the Great Alaska Earthquake of 1964," B. S. S. A., V58 No. 3, June, 1968, pp. 1131-1168.

(3) Richter, Charles F., Elementary Seismology, W. H. Freeman, San Francisco, 1958.

(4) Utsu, Tokuji, "A Method for Determining the Value of b in a Formula log n = a-bM Showing the Magnitude-Frequency Relation for Earthquakes," Geophys. Bull. Hokkaido Univ., V13, 1965, pp. 99-103.

(5) Utsu, Tokuji, "A Statistical Significance Test of the Difference in b-Value Between Two Earthquake Groups," Jour. of Physics of the Earth, V14, No. 2, 1966, pp. 37-40.

(6) Utsu, Tokuji, "Some Problems of the Frequency Distribution of Earthquakes in Respect to Magnitude," Geophys. Bull. Hokkaido Univ., V17, pp. 85-112; V18, pp. 53-69, 1967.

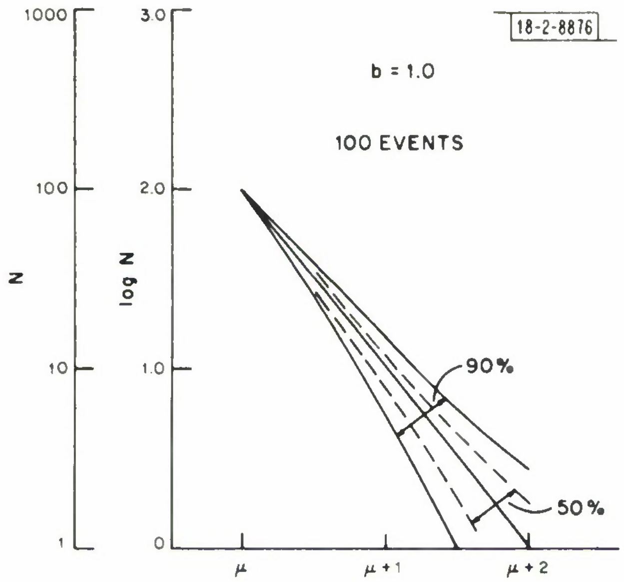


Fig. 1. Confidence intervals of cumulative histogram seismicity law given that 100 events occurred.

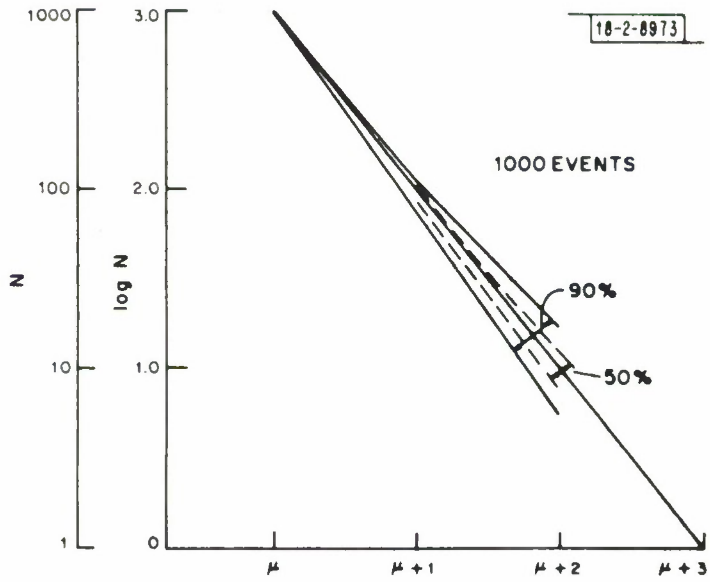
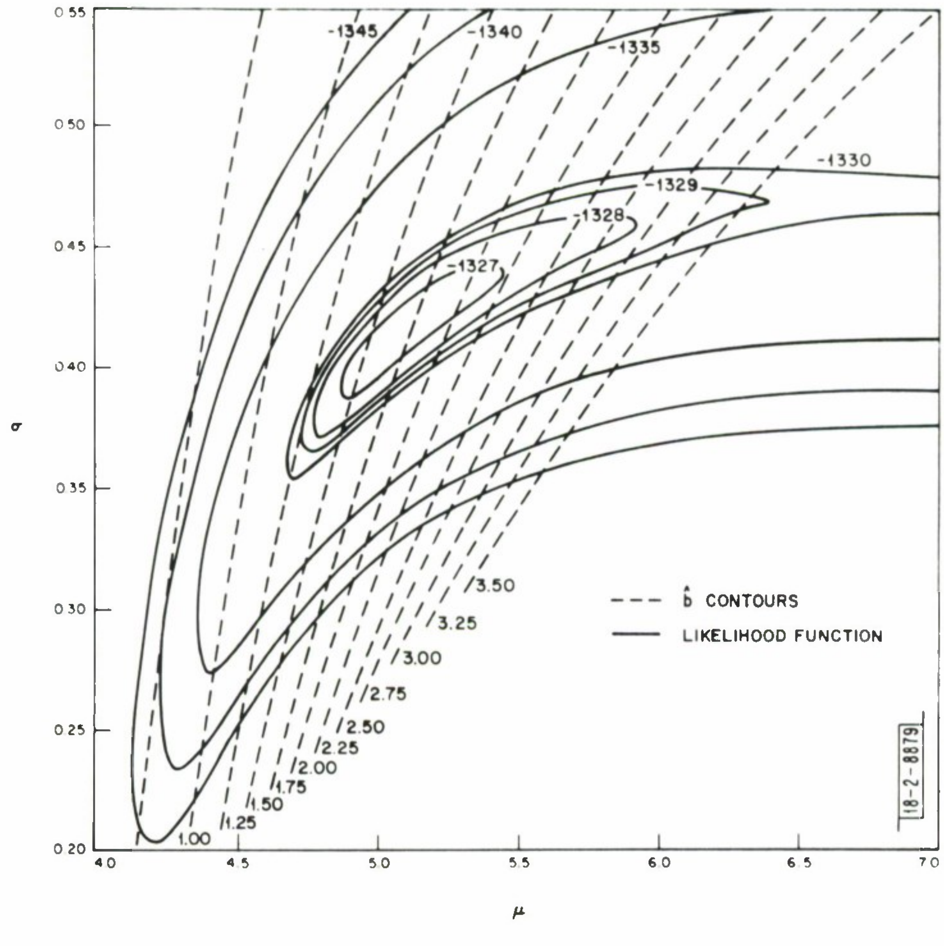
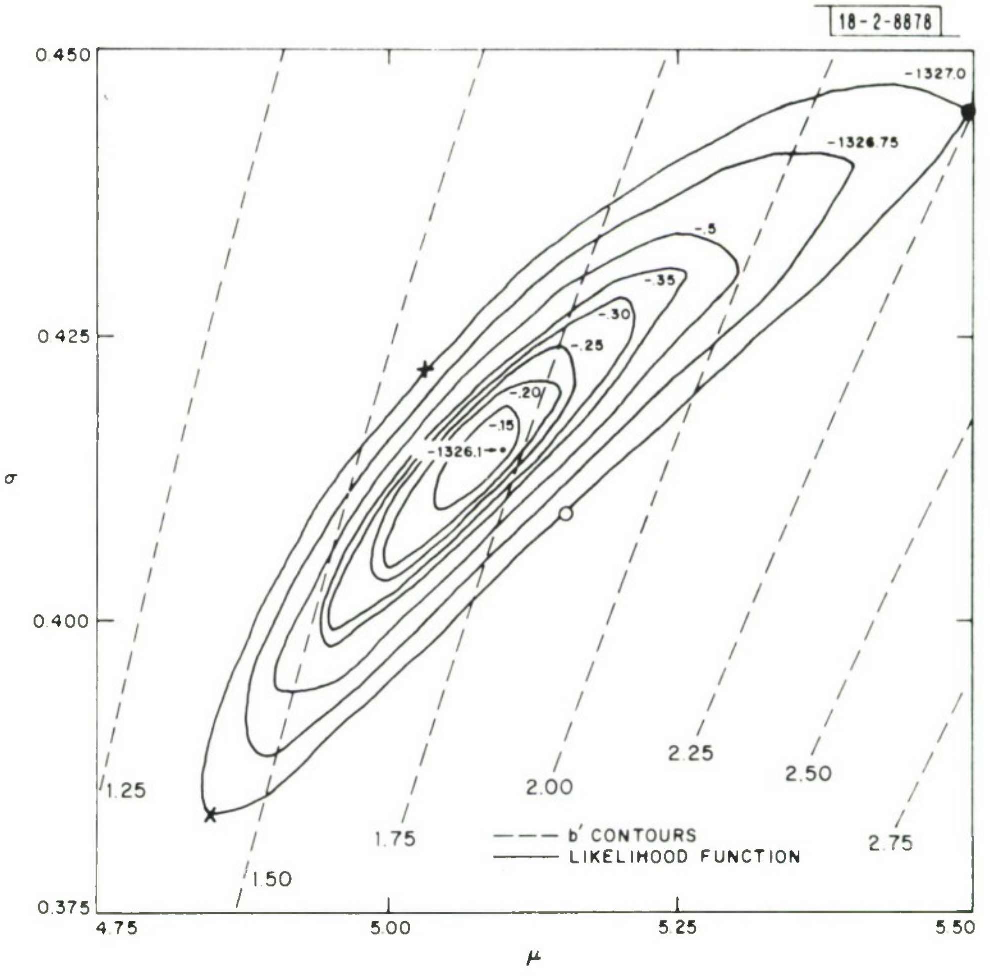


Fig. 2. Confidence intervals of cumulative histogram seismicity law given that 1000 events occurred.



(a)

Fig. 3. Likelihood function contours and contours of estimates of the slope of the seismicity curve.



(b)

Fig. 3. Continued.

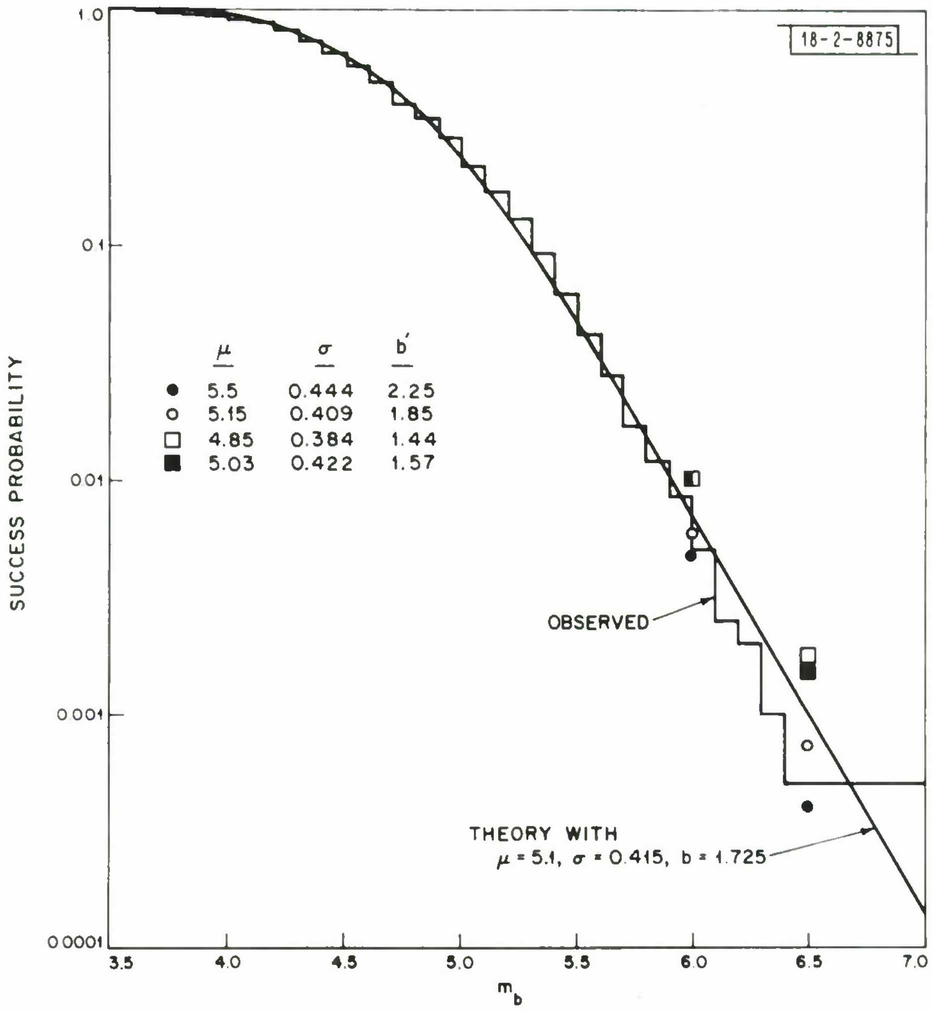


Fig. 4. Observed (2000 U. S. C. and G. S. events) and theoretical cumulative histograms.

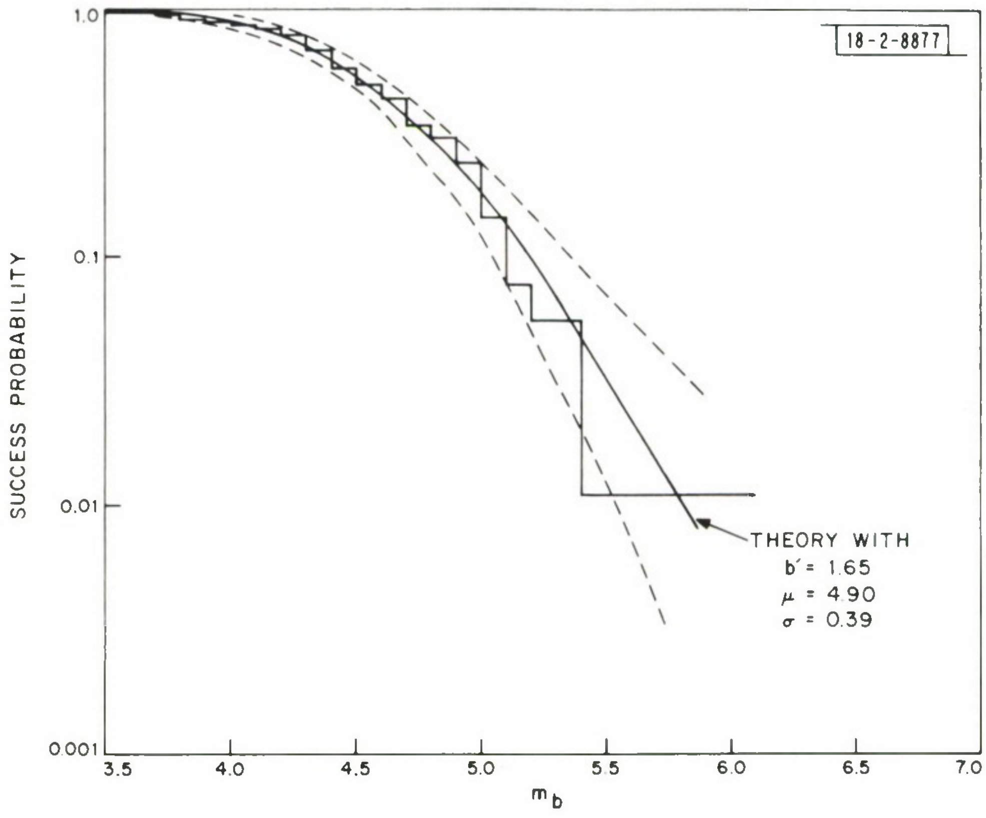


Fig. 5. Observed (90 U.S.C. and G.S. events) and theoretical cumulative histograms and 90% confidence intervals.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Lincoln Laboratory, M.I.T.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP None	
3. REPORT TITLE Estimation of Seismicity and Network Detection Capability		
4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* Technical Note		
5. AUTHOR(S) *(Last name, first name, initial)* Kelly, Edward J. and Lacoss, Richard T.		
6. REPORT DATE 16 September 1969	7a. TOTAL NO. OF PAGES 46	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. AF 19(628)-5167 b. PROJECT NO. ARPA Order 512 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Note 1969-41 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* ESD-TN-69-250	
10. AVAILABILITY/LIMITATION NOTICES This document has been approved for public release and sale; its distribution is unlimited.		
11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency, Department of Defense	

13. ABSTRACT

The problem of estimating seismicity and the performance of a system which detects earthquakes is formulated in such a way that maximum likelihood estimation can be applied. The mean number of earthquakes which occur in a fixed time interval is assumed to be of the form exp [a-bm] where m is magnitude and a and b are constants. The probability of detection as a function of m is taken to be an error function with mean and variance μ and σ. Procedures to obtain maximum likelihood estimates of a, b, μ and σ are derived, discussed, and applied to experimental data to check the relevance of the theoretical development.

14. KEY WORDS

seismic discrimination
earthquakes